Faculty of Engineering and Information Technology
University of Technology, Sydney

Coupled Behavior Informatics: Modeling, Analysis and Learning

A thesis submitted in partial fulfillment of
the requirements for the degree of
Doctor of Philosophy

by

Can Wang

October 2013

CERTIFICATE OF AUTHORSHIP/ORIGINALITY

I certify that the work in this thesis has not previously been submitted for a degree nor has it been submitted as part of requirements for a degree except as fully acknowledged within the text.

I also certify that the thesis has been written by me. Any help that I have received in my research work and the preparation of the thesis itself has been acknowledged. In addition, I certify that all information sources and literature used are indicated in the thesis.

Signature of Candidate

Can Wang

Acknowledgments

First and foremost, I would like to express the deepest appreciation to my supervisor, Prof. Longbing Cao, for providing me with continuous support throughout my PhD study and research. Without his professional guidance and persistent help, this dissertation would not have been possible.

Thanks to my co-workers, Jinjiu Li and Wei Wei, for being so supportive especially when I was struggling through hard times. We three enrolled in the same semester at a close time, through all these years, we shared our joy of achievements and the anxiety of struggling. Luckily, we have got over the hard times and shall step into our new period of life.

I am also grateful to my colleagues Xin Cheng and Chunming Liu, for their hard working in our collaborated papers. Thanks to all other colleagues in AAI, their warm help and insightful suggestions have definitely helped me. Also I would like to thank all the support staffs in AAI.

I would like to in particular thank my boyfriend Zhong She, without his dedicated support, selfless help and endless love, I can not survive and outlast the tough times during my PhD period. I would like to attribute my productive phase, i.e. the last two years, to him without a hesitate.

Last but not the least, I would like to thank my parents. Without their encouragement, finishing this dissertation would be impossible.

Can Wang
June 2013 @ UTS

Contents

Certificate i
Acknowledgment ii
List of Figures vii
List of Tables x
Abstract xii

Chapter 1 Introduction 1
1.1 Background 1
1.1.1 Behavior 1
1.1.2 Coupled Behaviors 3
1.1.3 Modeling, Analysis and Learning 9
1.2 Limitations and Challenges 11
1.3 Research Issues and Objectives 15
1.4 Research Contributions 17
1.4.1 Coupled Behavior Formalization and Verification . . . 18
1.4.2 Numerical Coupled Behavior Analysis 18
1.4.3 Categorical Coupled Behavior Analysis 19
1.4.4 Coupled Behavior Ensemble Learning 20
1.5 Thesis Organization 21

Chapter 2 Literature Review 24
2.1 Behavior Modeling and Analysis 25
2.1.1 Demographic and Transactional Perspective 26
2.1.2 Behavioral Perspective 27

2.1.3 Related Techniques 29
2.2 Numerical Behavior Analysis 31
2.3 Categorical Behavior Analysis 32
2.3.1 Nominal Similarity in Unsupervised Learning 33
2.3.2 Categorical Clustering 35
2.4 Behavior Application: Clustering Ensemble Learning 36
2.4.1 Process of Clustering Ensemble 36
2.4.2 Problems on Consensus Functions 37
2.4.3 Other Related Issues 39
2.5 Summary 39

Chapter 3 Formalization and Verification of Group Behavior Interactions 42
3.1 Background and Overview 43
3.2 Coupled Behavior Modeling Framework 47
3.2.1 Modeling and Checking System 47
3.2.2 Case Study Description 50
3.3 Behavior Descriptor 51
3.3.1 Behavior Visual Descriptor 51
3.3.2 Behavior Formal Descriptor 57
3.4 Behavior Aggregator 61
3.4.1 Intra-coupled Aggregation 62
3.4.2 Inter-coupled Aggregation 64
3.4.3 Combined Aggregation 66
3.5 Case Study: Representing Complex Behavior Systems 68
3.6 Behavior Constraint Indicator 72
3.7 Behavior Checker 75
3.8 Behavior Model Refiner and Exporter 79
3.9 Summary 80

Chapter 4 Numerical Coupled Behavior Analysis 83
4.1 Background and Overview 84

4.2 Framework of Coupled Attribute Analysis on Numerical Data 88
4.3 Coupled Interactions of Attributes 89
4.4 Coupled Representation for Objects 93
4.5 Empirical Study 96
4.5.1 Parameter Estimation 97
4.5.2 Learning Applications 98
4.6 Summary 103

Chapter 5 Categorical Coupled Behavior Analysis 104
5.1 Background and Overview 105
5.2 Preliminary Definitions 109
5.3 Framework of the Coupled Attribute Similarity Analysis . . . 112
5.4 Coupled Attribute Similarity 113
5.4.1 Intra-coupled Interaction 114
5.4.2 Inter-coupled Interaction 115
5.4.3 Coupled Interaction 120
5.5 Theoretical Analysis 121
5.6 Coupled Similarity Algorithm 123
5.7 Experiments and Evaluation 124
5.7.1 Coupled Similarity Comparison 127
5.7.2 Learning Applications 132
5.8 Summary 144
5.9 Appendix: Proofs 144

Chapter 6 Coupled Behavior Ensemble Learning 158
6.1 Background and Overview 160
6.2 Preliminary Definitions 166
6.3 The Coupled Framework of Clustering Ensembles 168
6.4 Coupling relationships in *CCE* 170
6.4.1 Coupling of Base Clusterings 170
6.4.2 Coupling of Objects 174
6.5 Coupled Consensus Function in *CCE* 178

6.5.1 Clustering-based Coupling 178
6.5.2 Object-based Coupling 180
6.5.3 Cluster-based Coupling 182
6.5.4 Miscellaneous Issues 183
6.6 Algorithm and Analysis 184
6.7 Empirical Study 187
6.7.1 Data Sets 188
6.7.2 Baseline Approaches and Parameters 188
6.7.3 Experimental Results 191
6.7.4 Data Characteristics and Performance 200
6.8 Summary 205

Chapter 7 Integrated Understanding with Discussions 207
7.1 Consolidated Understanding 207
7.2 Multi-level Couplings 210
7.3 Preliminary Coupled Behavior Algebra 213
7.4 Discussions with Open Issues 218
7.4.1 Qualitative Coupled Behaviors 218
7.4.2 Quantitative Coupled Behaviors 220
7.4.3 Integrated Coupled Behaviors 224

Chapter 8 Conclusion and Future Work 226
8.1 Conclusion 226
8.2 Future Work 229
8.2.1 Qualitative Behavior Analytics 229
8.2.2 Quantitative Behavior Analytics 230
8.2.3 Consolidated Understanding 232

Chapter A Appendix: List of Publications 233

Chapter B Appendix: List of Symbols 236

Bibliography 239

List of Figures

1.1 Qualitative coupled behaviors: Robocup soccer competition. . 6
1.2 Quantitative coupled behaviors: clustering ensemble. 7
1.3 Research issues. 15
1.4 The profile of research work in this thesis. 22

2.1 The framework of literature review. 25

3.1 General framework of the *OntoB* system. 48
3.2 A case-based multi-robot system with n robots and k retrievers. 50
3.3 The behavior ontology that involves coupling (i.e. coupled behaviors). 52
3.4 Relationships between the qualitative coupled behaviors: *Instance Of* (“$-\cdot\rightarrow$”), *Subclass Of* (”$\longrightarrow$”), *Unit Feature* (“$\dashrightarrow$”). 54
3.5 Transition system models $TS(\mathbb{B}(RC_p))$ and $TS(\mathbb{B}(Ord_q))$. . . 63
3.6 Group behavior representation and verification procedures. . . 69
3.7 A simplified case-based multi-robot system. 70
3.8 Ontology-based Qualitative Coupled Behavior Modeling and Checking ($OntoB$). 77
3.9 A scenario in a soccer game that causes deadlock. 78

4.1 A framework of the coupled attribute analysis. 88
4.2 The performance of L on six data sets: the average accuracy with $\pm$ sample standard deviation error bars. 97
4.3 Data structure index comparisons on nine data sets. 99

4.4 Data classification comparisons on six data sets: the average values with $\pm$ sample standard deviation error bars. 102

5.1 A framework of coupled attribute similarity analysis, where $\dashleftarrow\dashrightarrow$ indicates intra-coupling and $\longleftrightarrow$ refers to inter-coupling. 113
5.2 Complexity on individual attributes. 128
5.3 Scalability on n and R respectively. 131
5.4 Data structure index comparison. 134
5.5 Clustering evaluation on six data sets. 137
5.6 Clusterings for conferences/journals. 142
5.7 Comparisons on classification. 143

6.1 Four possible base clusterings of 12 data objects into two clusters; different partitions use different sets of labels. 161
6.2 A graphical representation of the coupling relationship between base clusterings, where a circle denotes an object, a rounded rectangle represents an cluster, and an edge exists if an object belongs to a cluster. 163
6.3 A coupled framework of clustering ensembles (CCE), where $\dashleftarrow\dashrightarrow$ indicates the intra-coupling and $\longleftrightarrow$ refers to the inter-coupling. 170
6.4 An example of the coupled similarity for cluster labels α and β, where $\dashleftarrow\dashrightarrow$ indicates the intra-coupling and $\longleftrightarrow$ refers to the inter-coupling, with the value along each line being the corresponding similarity. 174
6.5 Percentage of pairwise similarity $\in (0, 1)$ between objects among all the similarity values. 192
6.6 Clustering-based comparisons on AC, NMI and CSI. 195
6.7 Object-based comparisons on AC, NMI and CSI. 197
6.8 Quality of base clusterings and the AC performance gain for the results in Figure 6.6. 200

6.9 Quality of base clusterings and the NMI performance gain for the results in Figure 6.7. 202
6.10 Consistency of base clusterings and the AC, NMI performance gains for the results in Table 6.7. 203

7.1 A coupled behavior information table. 209
7.2 Hierarchical structure of couplings. 211

8.1 Prospects for the coupled behavior informatics 227

List of Tables

3.1 List of Main Notations in Chapter 3 49

4.1 A Fragment Example of Iris Data Set 85
4.2 The Extended Information Table of Iris Data Set 90
4.3 The Coupled Representation of Iris Data Set 95
4.4 Description of Data Sets in Chapter 4 96
4.5 Clustering Comparisons on Six Data Sets with ± Sample Standard Deviation 100

5.1 An Instance of the Movie Database 107
5.2 An Example of Information Table 110
5.3 List of Main Notations in Chapter 5 112
5.4 Example of Computing Similarity Using *IRSP* 117
5.5 Computing Similarity Using *IRSU* 118
5.6 Computing Similarity Using *IRSJ* 119
5.7 Computing Similarity Using *IRSI* 119
5.8 Time Cost of *ICP* 122
5.9 Computational Complexity for *CASV* 123
5.10 Description of Data Sets in Chapter 5 126
5.11 Complexity Comparison on All Attributes 129
5.12 *CROCK* vs *ROCK* on UCI Data Sets 138
5.13 Clustering Qualities for First Author 140

6.1 An Example of Base Clusterings 167

6.2 List of Main Notations in Chapter 6 168
6.3 List of Abbreviations in Chapter 6 169
6.4 An Example of Neighborhood Domain for Object 178
6.5 Description of Data Sets in Chapter 6 187
6.6 Evaluation Measures on Base Clusterings 191
6.7 Cluster-based Comparisons on AC, NMI and CSI 199

Abstract

Behavior refers to the action or property of an actor, entity or otherwise, to situations or stimuli in its environment. The in-depth analysis of behavior has been increasingly recognized as a crucial means for understanding and disclosing interior driving forces and intrinsic cause-effects on business and social applications, including web community analysis, counter-terrorism, fraud detection and customer relationship management, etc. Currently, behavior modeling and analysis have been extensively investigated by researchers in different disciplines, e.g. psychology, economics, mathematics, engineering and information science. From those diverse perspectives, there are widespread and long-standing explorations on behavior studies, such as behavior recognition, reasoning about action, interactive process modeling, multivariate time series analysis, and outlier mining of trading behaviors.

All the above emerging methods however suffer from the following common issues and problems to different extents: (1) Existing behavior modeling approaches have too many styles and forms according to distinct situations, which is troublesome for cross-discipline researchers to follow. (2) Traditional behavior analysis relies on implicit behavior and explicit business appearance, often leading to ineffective and limited understanding on business and social activities. (3) Complex coupling relationships between behaviors are often ignored or only weakly addressed, which fails to provide a complete understanding of the underlying problems and their comprehensive solutions. (4) Current research usually overlooks the checking of behavior interactions, which weakens the soundness and robustness of models built for complex behavior applications. (5) Most of the classic mining and learning algorithms follow the fundamental assumption of independent and identical distribution (i.e. IIDness), but this is too strong to match the reality and complexities in practical applications.

With the deepening and widening of social/business intelligences and their networking, the concept of behavior is in great demand to be consolidated and formalized to deeply scrutinize the native behavior intention, lifecycle and impact on complex problems and business issues. In the real-world applications, group behavior interactions (i.e. coupled behaviors) are widely seen in natural, social and artificial behavior-related problems. The verification of behavior modeling is further desired to assure the reliability and stability. In addition, complex behavior and social applications often exhibit strong explicit or implicit coupling relationships both between their entities and properties. They can not be abstracted or weakened to the extent of satisfying the IIDness assumption. These characteristics greatly challenge the current behavior-related analysis approaches. Moreover, it is also very difficult to model, analyze and check behaviors coupled with one another due to the complexity from data, domain, context and impact perspectives.

Based on the above research limitations and challenges, this thesis reports state-of-the-art advances and our research innovations in modeling, analyzing and learning coupled behaviors, which constitute the coupled behavior informatics. Coupled behaviors are categorized as qualitative coupled behaviors and quantitative coupled behaviors, depending on whether the behavior involved is qualified by actions or quantified by properties.

In terms of the qualitative coupled behavior modeling and analysis, we propose an Ontology-based Qualitative Coupled Behavior Modeling and Checking (*OntoB* for short) system to explicitly represent and verify complex behavior relationships, aggregations and constraints. The effectiveness of *OntoB* system in modeling multi-robot behaviors and their interactions in the Robocup soccer competition game has been demonstrated.

With regard to the quantitative coupled behavior analysis and learning, we carry out explorations on three tasks below. They are under the non-IIDness assumption of entities or properties or both of them, which caters for the intrinsic essence of real-world problems and applications.

For *numerical coupled behavior analysis*, we introduce a framework to address the comprehensive dependency among continuous properties. Substantial experiments show that the coupled representation can effectively model the global couplings of numerical properties and outperforms the traditional way. For *categorical coupled behavior analysis*, we present an efficient data-driven similarity learning approach that generates a coupled property similarity measure for nominal entities. Intensive empirical studies witness that the coupled property similarity can appropriately quantify the intrinsic and global interactions within and between categorical properties for especially large-scale behavior data. For *coupled behavior ensemble learning*, we explicate the couplings between methods and between entities in the application of clustering ensembles, and put forward a framework for coupled clustering ensembles (*CCE*). The *CCE* is experimentally exhibited to capture the implicit relationships of base clusterings and entities with higher clustering accuracy, stability and robustness, compared to existing techniques. All these models and frameworks are supported by statistical analysis.

Finally, we provide a consolidated understanding of coupled behaviors by summarizing the qualitative and quantitative aspects, extract the multi-level couplings embedded in them, and then formalize a coupled behavior algebra at its preliminary stage. Many open research issues and opportunities related to our proposed approaches and this novel algebra are discussed accordingly.

Under varying backgrounds and scenarios, our proposed systems, algorithms and frameworks for the coupled behavior informatics are evidenced to outperform state-of-the-art methods via theoretical analysis or empirical studies or both of them. All these outcomes have been accepted by top conferences, and the follow-up work has also been recognized. Therefore, coupled behavior informatics is a promising though wholly new research topic with lots of attractive opportunities for further exploration and development.